## Correlation, Correlation analysis

Correlation analysis is an investigation of the measure of statistical association among random variables based on samples. Widely used measures include the *linear correlation coefficient* (also called *product-moment correlation coefficient*  or *Pearson’s correlation coefficient*), and such non-paraemtric measures as *Spearman rank-order correlation coefficient*  and *Kendall’s tau*. When the data are nonlinear, non-parametric correlation is generally considered to be more robust than linear correlation.

**Latin Hypercube Sampling**

**Monte Carlo Analysis, Monte Carlo Simulation**

Monte Carlo Analysis uses statistical sampling techniques in obtaining a probalistic approximation to the solution of a mathematical equation or model. In the contect of Monte Carlo Analysis, simulation is the process of approximating the output of a model through repetitive randm application of a model’s algorithm.

**Sensitivity, Sensitivity analysis**

Sensitivity generally refers to the variation in output of a mathematical model with respect to changes in the values of the model’s input or parameters. In a broader sense, sensitivity can refer to how conclusions may change if models, data, or assessment assumptions are changed.

**Uncertainty**

Uncertainty refers to *lack of knowledge* about specific factors, parameters or models. Uncertainty includes *parameter uncertainty* (measurement errors, sampling errors, systematic errors), *model uncertainty* (uncertainty due to necessary simplification of real-world processes, mis-specification of teh model structure, model misuse, use of inappropriate surrogate variables), and *scenario uncertainty* (descriptive errors, aggregation errors, errors in professional judgment, incomplete analysis).

**Variability**

Variability refers to observed differences attributable to *true heterogeneity* or diversity in a population or parameter. Sources of variability are the result of natural random processes and stem from environmental, lifestyle, and genetic differences among humans. Examples include human physiological variation (e.g. natural variation in bodyweight, height, breathing rates, …).

# Guiding principles for Monte Carlo Analysis

Conduct prelimenary sensitivity analyses or numerical experiments to identify model structures, exposure pathways, and model input assumptions and parameters that make important contributions to overall variability and/or uncertainty.

Dependencies or correlations between model parameters also may have a significant influence on the outcome of the analysis. Those dependencies or correlations identified must be accounted for in later analyses.

**Methods should investigate the numerical stability of the moments and the tails of the distributions**

Numerical stability referes to observed numerical changes in characteristics (i.e. mean, variance, percentiles) of the Monte Carlo simulation output distribution as the number of simulations increases. Random number seed should always be recorded. Ideally, Monte Carlo simulations should be repeated using several non-overlapping subsequences to check for stability and repeatability.

The analysi should clearly disclose what set of uncertainties the analysis attempts to represent and what does it not.

**Calculate and present point estimates**

Traditional deterministic (point) estimates should be calculated using established protocols.

**Gibbs Sampling**

Gibbs sampling is applicable when the joint distribution is not known explicitly or is difficult to sample from directly, but the [conditional distribution](https://en.wikipedia.org/wiki/Conditional_distribution) of each variable is known and is easy (or at least, easier) to sample from. The Gibbs sampling algorithm generates an instance from the distribution of each variable in turn, conditional on the current values of the other variables. It can be shown (see, for example, Gelman et al. 1995) that the sequence of samples constitutes a [Markov chain](https://en.wikipedia.org/wiki/Markov_chain), and the stationary distribution of that Markov chain is just the sought-after joint distribution.

Gibbs sampling is particularly well-adapted to sampling the [posterior distribution](https://en.wikipedia.org/wiki/Posterior_probability) of a [Bayesian network](https://en.wikipedia.org/wiki/Bayesian_network), since Bayesian networks are typically specified as a collection of conditional distributions.

**Metropolis Hasting**

In [statistics](https://en.wikipedia.org/wiki/Statistics) and in [statistical physics](https://en.wikipedia.org/wiki/Statistical_physics), the Metropolis–Hastings algorithm is a [Markov chain Monte Carlo](https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo) (MCMC) method for obtaining a sequence of [random samples](https://en.wikipedia.org/wiki/Pseudo-random_number_sampling) from a [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution) for which direct sampling is difficult. This sequence can be used to approximate the distribution (i.e., to generate a [histogram](https://en.wikipedia.org/wiki/Histogram)), or to [compute an integral](https://en.wikipedia.org/wiki/Monte_Carlo_integration) (such as an [expected value](https://en.wikipedia.org/wiki/Expected_value)). Metropolis–Hastings and other MCMC algorithms are generally used for sampling from multi-dimensional distributions, especially when the number of dimensions is high. For single-dimensional distributions, other methods are usually available (e.g. [adaptive rejection sampling](https://en.wikipedia.org/wiki/Adaptive_rejection_sampling)) that can directly return independent samples from the distribution, and are free from the problem of auto-correlated samples that is inherent in MCMC methods.